
NLSY79 CODEBOOK SUPPLEMENT

MAIN FILE

1979-1998

Table of Contents

Attachments

NLSY79 Attachment 3: 1970 & 1980 Census Industrial & Occupational Classification Codes and 1977 Department of Defense Enlisted Occupational Codes.....	1
NLSY79 Attachment 4: Fields of Study in College	51
NLSY79 Attachment 5: Index of Labor Unions and Employee Associations	59
NLSY79 Attachment 6: Other Kinds of Training Codes.....	69
NLSY79 Attachment 7: Other Certificate Codes	73
NLSY79 Attachment 8: Health Codes.....	79
NLSY79 Attachment 100: Geographic Regions	87
NLSY79 Attachment 101: Country Codes.....	91
NLSY79 Attachment 102: Federal Information Processing Standards Publication 5-1 June 15, 1970.....	97
NLSY79 Attachment 103: Religion Codes.....	103

Appendices

NLSY79 Appendix 1: Employment Status Recode Variables 1979-1998	109
NLSY79 Appendix 2: Total Net Family Income Variable Creation: 1979-1998.....	183
NLSY79 Appendix 3: Job Satisfaction Measures	271
NLSY79 Appendix 4: Job Characteristic Index 1979-1982	275
NLSY79 Appendix 5: Supplemental Fertility File Variables (1998)	279
NLSY79 Appendix 6: Urban-Rural and SMSA-Central City Variables	287
NLSY79 Appendix 7: Unemployment Rate	291
NLSY79 Appendix 8: Highest Grade Completed & Enrollment Status Variable Creation: 1990- 1998.....	295
NLSY79 Appendix 9: Linking Employers Through Survey Years.....	331
NLSY79 Appendix 11: NLSY79 Round 12 (1990) Survey Administration Methods	337
NLSY79 Appendix 12: Most Important Job Learning Activities (1993-94)	341
NLSY79 Appendix 14: 1993-1998 Instrument Rosters.....	367
NLSY79 Appendix 15: Reciprocity Event Histories	383
NLSY79 Appendix 16: 1994 Recall Experiment	399
NLSY79 Appendix 17: ISIS Data	405

NLSY79 APPENDIX 13

INTRODUCTION TO 1993 THROUGH 1998

CAPI QUESTIONNAIRES AND

CODEBOOKS

Introduction to 1993 CAPI Questionnaire and Codebook

Round 15 (1993) marks the first round of the National Longitudinal Survey of Youth to be administered entirely in CAPI (Computer-Assisted Personal Interviewing) mode. Wherever possible, comparability has been maintained between the 1993 data documentation and that of previous survey years. However, changes in technology and certain data collection procedures have resulted in some important differences. A number of these changes are discussed below.

I. NEW DOCUMENTATION ITEMS AND CHANGES IN EXISTING DOCUMENTATION

Questionnaire

An introductory document entitled *1993 NLSY79 Questionnaire* assists the user in understanding the format and content of the CAPI questionnaire. This document precedes the questionnaire, which is included in this *Codebook Supplement*.

Glossary of Save Array Names

A glossary of save array names accompanies the questionnaire. See Section II below and *1993 NLSY79 Questionnaire* introduction for a discussion of save arrays. These names are attached to data storage places which help the CAPI instrument to function. The save array names appear in numerous question texts in both the codebook and the questionnaire. The *Glossary of Save Array Names* defines the information stored in each save array as it is used during the course of the survey.

1993 Instrument Rosters

A listing of rosters, such as the Household Roster, has been added to this *Codebook Supplement*. This document is entitled *1993 Instrument Rosters*. This listing includes the contents of the rosters, the applicable save array names attached to the rosters that might be encountered in the codebook and/or the questionnaire, and the variable reference number assigned to each piece of data in the codebook. Further detail about the rosters are provided in Section II.

Question by Question Specification (Q by Q)

The 1993 Q by Q is different in both appearance and content from those distributed during earlier rounds. Previously, the Q by Q resembled an actual questionnaire, with the question-by-question specifications for the interviewer included. In essence, the QxQs could be used as a questionnaire. The 1993 Q by Q however, is a combination of the electronic "help screens" that were available to the interviewer for specific questions, and a more detailed hard copy outline of the sections of the questionnaire and question intent. The electronic "help screens" contain listings of the questions for which they were used during the interview. The 1993 Q by Q should be used in conjunction with the codebook or the questionnaire rather than as an independent document.

II. NEW TERMS

Save Arrays and "Text Fills"

Save arrays are reserved fields into which data can be placed, stored and accessed throughout the questionnaire. Each save array field is assigned a distinct name, often with an index number appended when the same piece of information is stored for a set of subjects. For instance, the save array "lintdate" contains the date of last interview for the respondent. The save arrays "child.name(1)" through "child.name(9)" contain the names of the first through the ninth biological child of the respondent. Throughout the survey, the last interview date and/or the name of the respondent's first child can be accessed by invoking the save array names "lintdate" and "child.name(1)" respectively.

Users will encounter these save array names both in the codebook and in the questionnaire. Sometimes data in a save array is accessed and used to govern a skip or perform a calculation. At other times, a save array name may be part of a question text. In this case, the content of the specific save array content becomes part of the text of the question, and is read as such. This is referred to as a "text substitution" or "text fill". For instance, a question text reading, "How often does (child.name(1)) see (his/her) (mom/dad)?" would appear to the interviewer with the appropriate name replacing the save array "child.name(1)", the appropriate gender replacing the save array "his/her" and the appropriate parent (mother or father) replacing the save array "mom/dad". Such alternative wordings for text substitutions have been automated to a great extent with the advent of CAPI. In previous PAPI (Paper and Pencil Interviewing) questionnaires, the interviewer was required to make these choices as s/he read the question. (See *1993 NLSY79 Questionnaire* for more discussion of save arrays and text fills.)

Loops

As in PAPI instruments, certain sequences of questions in the CAPI instrument are repeated a number of times. For instance, some sets of questions are repeated up to 20 times in the Fertility Section (Section 9) of the 1993 CAPI questionnaire, for each of up to 20 children. Similarly, a sequence of questions is repeated ten times in the On Jobs Section, (Section 5), allowing up to ten new employers to be reported. Question names that include a ".1", ".2", ".3", etc. at the end, belong to these repeating sequences of questions. Each repetition of the sequence of questions is referred to as a "loop". Taking the Fertility Section as an example, the sequence of questions asking about the first child is referred to as the first "loop". The sequence asking about the second child is referred to as the second "loop", and so on. While the codebook generally includes more than one loop in a series (because more than one may contain valid data), the questionnaire includes only the **first loop** in each set of loops. (See *1993 NLSY79 Questionnaire* for further information.)

Dummy Records

A "dummy record" is a type of record or variable containing text that functions either as an explanatory statement or a transition. This text is read to the respondent, or less frequently, serves as an instruction to the interviewer. No information is entered when a dummy record appears on the screen. These screens are presented in questionnaire, and hard-copy codebook. These dummy records however, are **not included** on the CD-ROM. When a dummy record is encountered in the codebook, it is present only to represent the flow of the questionnaire and generally appears as text preceding a variable with valid data. Although these Dummy Records have been assigned reference numbers, users should not attempt to use them as data items. These variables are marked by the text "(DUMMY)" at the end of the KWIC TITLE in the codebook.

Rosters and Roster Edit Records

A "roster" is a listing of one or more items of information pertaining to a specific set of subjects, such as the biological children of the respondent, or members of the respondent's household. For instance, the CHILD roster contains a variety of information pertaining to each biological child of the respondent, (e.g., name, gender, birthdate, etc.). By using the roster format, an inventory of information can be gathered, tagged to a specific subject, carried along through the interview and accessed when necessary. This format also allows for these items of information to be presented to the interviewer at any time.

A "roster edit" is a type of record or variable in which the interviewer is presented with a roster or listing of information pertaining to a given set of subjects. The roster is then used by the interviewer to either verify information concerning this list of subjects, or to choose one of the subjects from the list, as indicated by the respondent. For example, after the respondent has reported all biological

children, the interviewer is presented with the entire CHILD roster so that the s/he can verify with the respondent, the children listed. The interviewer is also presented with a modified roster of biological child information from which to choose the child (or children) for whom corrective information is required. These variables are marked by the text "(RE)" at the end of the KWIC TITLE in the codebook or by the text "ROSTER EDIT" in a note contained in the codeblock.

III. CHANGES IN CODING CONVENTIONS

Hard and Soft Range Restrictions

The "Hard Minimum", "Hard Maximum", "Soft Minimum" and "Soft Maximum" specifications control the allowable range of values that can be entered for a given question. These fields are only active when a question calls for the entry of a time date or amount. Questions that require the interviewer to select one response or to select all that apply do not contain this field, as the range limits are implicit in the distribution code block and are thereby enforced.

Hard minima and maxima are absolute limits that an interviewer- or respondent-generated answer must obey. Entry of values outside the hard range is not allowed. In such cases the interviewer is instructed to enter the maximum or minimum allowable value, as appropriate, enter the actual response in the comment field, and "flag" the case for central office checking. Soft minima and maxima are nested within the hard range. When a response falls outside the soft range but inside the hard range, the computer beeps and asks the interviewer to either confirm or change the response. These range checks are analogous to what took place in the central office coding shops, except that outliers are immediately brought to the interviewer's attention so corrective action can be taken as necessary while the respondent is present. As one would expect, the number of callbacks to respondents to collect missing or seemingly incorrect data has fallen to virtually zero.

In some cases, the hard ranges are themselves determined by a variable. For example, event history data collection imposes certain logical limits on dates that are acceptable in subsequent questions. Whenever a range restriction is contingent on data collected during the current, or a previous, interview, this is indicated in the hard minimum or maximum field. When a question requires that the latest date that can be entered is the current interview date, the Hard Maximum field will contain the notation: Hard Max: Month[%curdate%] Day[%curdate%] Year[%curdate%]. This indicates that the various fields of the date are restricted so that no date later than that date can be entered. This is a powerful tool for event histories (among other types of data sequences) and is used extensively in the instrument. (See also the *Glossary of Save Array Names* and *1993 NLSY79 Questionnaire*, for more information.)

Family Unit on Household Enumeration

In each survey year, a "family unit" code has been assigned to each member of the respondent's household. This code identified whether the household member belonged to the respondent's family unit or to other family units, possibly with other household members. Beginning with the 1993 survey, this family unit code has been collapsed into three codes. A code of "1" retains its traditional interpretation - a blood, marriage or adoptive relationship to the respondent. A new code of "2" represents a same sex or opposite sex partner and any household members related to the partner by blood or marriage. A code of "9" represents all other household members, with no familial relationship to either the respondent or his/her partner. This change was made in part because of the questionable reliability of these variables in past years, specifically as they were assigned to household members not in the respondent's family unit.

IV. CHANGES IN PLACEMENT OF DATA AND/OR CODEBOOK ORDER

Employer Supplements

The Employer Supplement data are now placed in the codebook in the actual order of administration, directly following the CPS section (Section 6, in the 1993 instrument). Previously, Employer Supplement data had been placed after the main body of the survey, the information sheet data, and the household enumeration data.

1980 CPS Occupation and Industry Codes

In the 1993 CAPI data, the 1980 Occupation and Industry codes are contained in the Employer Supplement instead of in the CPS section, as has traditionally been the case. These variables were coded and entered independently of the actual interview data, but were placed within the structure of the interview data for release.

Employer ID From Previous Survey Year

Employer IDs from the previous survey year have been placed directly preceding the wage questions in the Employer Supplements instead of directly preceding the start dates at the beginning of the Employer Supplements. In earlier PAPI rounds, the previous employer ID number was an independent data item, entered into the Employer Supplement by the interviewer during the course of the interview. Thus, these ids were part of the flow of the instrument. In the CAPI instrument this id number simply carried through the instrument in association with the appropriate employer on the EMPLOYER roster.

V. CHANGES IN APPEARANCE OR PRESENTATION OF DATA

Some 1993 CAPI questions generate more than one variable. For example, some questions collect information about the date an event happened and generate three variables: month, day and year. Similarly, questions that ask the respondent to indicate which of several responses are appropriate, and to pick all responses that apply, can generate multiple variables. In this case, each possible outcome becomes a distinct variable and is coded as "selected", "not selected", or a "valid skip". When a question record generates multiple variables, those variables have decimals in the reference numbers. These two types of questions (date-entry and "code-all-that-apply") and the examples of the resulting variables, are discussed further below.

Date-Entry Questions

All date questions, whether full dates (month, day and year) or just month and year, are represented in the CAPI codebook with respectively, three or two variables. Each variable contains the same codeblock displaying the ranges and missing values for all elements of the date. A base reference number ending in ".00", is assigned to the first variable in the set. The same base reference number, ending in ".01" and ".02" if necessary, is assigned to the other elements of the date. Thus, if the month is assigned a reference number of R41454., (DATE LEFT ARMED FORCES BRANCH SERVING IN AT TIME OF LAST INT (SRVD ACTIVE):MO), the day and year would be assigned a reference number of R41454.01 and R41454.02 respectively. If the date is only a month and year, such as R41453. (DATE LEFT ARMED FORCES BRANCH SERVING IN AT TIME OF LAST INT (SERVED INACTIVE):MO), then the year would be assigned R41453.01.

Code-all-that-apply Questions

The NLSY79 includes questions that allow respondents to give multiple responses or code all responses that apply. These "code-all-that-apply" questions have traditionally been presented in the codebook as separate variables, with codes representing one specific substantive answer per variable,

or an "applies"/"does not apply" coding scheme for each possible response. In the CAPI codebook, as in past codebooks, each possible response constitutes its own variable. Unlike such codeblocks in previous years, frequencies for **all** possible responses are represented in each codeblock for each variable. However, each specific variable contains **only** the valid data for one specific possible response. Reference numbers are assigned in the same manner as described for date-entry questions. A base reference number is assigned to the first possible response, such as R41752. (METHOD OF SEEKING EMPLOYMENT PAST 4 WKS (UNEMPLOYED) CPS ITEM:1), with a decimal value being appended to that base number for each following possible response. In this example these numbers range from R41752.01 to R41752.08. (Possible responses range from doing nothing, to checking with public and/or private employment agencies, friends, checking in newspapers, etc.) Although the codeblock for R41752.01 represents the frequencies for all possible responses to the question, accessing the data for that specific variable will produce only the data for the response "CHECKED WITH PUBLIC EMPLOYMENT AGENCY".

Time Stamp Variables

In order to simplify the questionnaire and the public use data file, we have dropped a number of elapsed machine time variables. This detailed timing data may be useful to researchers interested in response times and how much thought the respondent may have given to particular questions or sets of questions. If you need such data, please notify CHRR and we will provide additional information.

VI. CHANGES IN DATA COLLECTION PROCEDURES WHICH AFFECT FORM AND/OR APPEARANCE OF NLSY79 1993 DATA

Machine-Generated Check Items

Many questions represented in the NLSY79 data are "interviewer check items". These are items used by the interviewer, to help determine the flow or skip patterns to follow in the questionnaire. In the codebook, these check items have traditionally looked like other questions that were asked of the respondent. In other words, they were often posed in the form of a question, which the interviewer simply asks and answers her/himself. Such items are still present in the CAPI instrument and codebook. However, a great number of them are now "machine checks". Previously reported information is checked by machine and computations that once were made by the interviewer are made automatically, causing the appropriate skip pattern to be executed without intervention by the interviewer. In an effort to maintain comparability with past data releases and to clarify the skip patterns present in the instrument, a large number of these machine checks appear in the codebook. However the texts of these checks are in machine language and equations. Wherever possible, documentation comparable to previous check items in the PAPI questionnaires has been used. In addition, comments have been included as a "translation", to clarify the purpose of each machine check. R41750. (INT CHECK - IS R CURRENTLY LOOKING FOR WORK? CPS ITEM), is one such variable. The text of the machine check is "([Q6-1]=3);", an equation that is evaluated by the machine. A comment reading "/* WAS R 'LOOKING FOR WORK' LAST WEEK (Q6-1 CODED '3') */" however explains the question to be determined by the machine at this question.

"Consolidated" Variables

An effort has been made in the 1979-1993 data release to maintain comparability with previous releases in terms of data presentation. Toward this end, some variables have been "consolidated" with other variables. This means that the answers to more than one variable or set of variables are contained in a single variable or set of variables. This has generally been done in cases where data in previous years was available from a single variable or set of variables, but is now collected in the CAPI instrument in more than one variable or set of variables.

In each case, the variables being consolidated are mutually exclusive with respect to substantive responses. In other words, if variable A, variable B and variable C are consolidated with each other, respondents will have given a response to only one of these - either variable A, or variable B or variable C. This allows the user to access one variable or a smaller set of variables, as may have been possible in previous rounds, instead of having to access a number of different variables to get the same information previously available in one.

These consolidated sets of variables are noted in the codebook documentation in the following manner. Variables that contain the responses from other variables have notes in the codebook documentation indicating that the variable is "CONSOLIDATED WITH Q#-#" and that it "INCLUDES RESPONSES FROM Q#-#". For instance, R41463. - R41463.02 (Q4-11B - DATE BEGAN SERVICE IN BRANCH OF MOST RECENT ARMED FORCES ENLISTMENT (SRVD ACTIVE)), contain notes indicating that this set of variables is "CONSOLIDATED WITH Q4-11A" and that each variable in the set "INCLUDES RESPONSES FROM Q4-11A". Conversely, R41462. - R41462.01 (Q4-11A, DATE BEGAN SERVICE IN BRANCH OF MOST RECENT ARMED FORCES ENLISTMENT (SRVD INACT)), contain notes indicating that this set of variables has been "CONSOLIDATED WITH Q4-11B", but that "RESPONSES INCLUDED IN Q4-11B". In this case, the month/day/year dates of enlistment in the active armed services have been consolidated with the month/year dates of enlistment in the inactive armed services. While both sets of variables are present in the questionnaire, users need only access R41463. - R41463.02 to get data for both variables. This was done to maintain comparability with the previous PAPI data releases, in which users could get information for both groups of respondents by accessing only one set of variables reflecting enlistment dates for both active and inactive armed forces.

Consolidation spares users from having to access a larger number of variables and use each separately or combine the responses themselves.

Changed Patterns and Formats for Data Collection

CAPI has led to a change in the formats in which some pieces of data are collected. Military dates are one example. For those enlisting in, or departing from, inactive military service since last interview, only the month and year of separation is collected, not the day. Those in active military service are asked the full date of separation or enlistment, including the day. The PAPI data contained the month and year, and then the day for those serving in the active forces or on active duty. The CAPI data however, contains two sets of date variables - a month and a year for those in the inactive forces and a month, a day and a year for those in the active forces or on active duty. While the data collected is identical to that in the PAPI survey, the format and presentation is slightly different. These variables have been consolidated into one set of variables in order to help maintain comparability with the PAPI data from past rounds. (See discussion on consolidated variables in the CAPI data above.)

Identification of CPS Employer (Current/Most Recent Employer)

The mechanism for identifying the CPS employer (i.e. the current or most recent employer), has changed substantially with the implementation of a CAPI instrument. In previous PAPI rounds, identification of the CPS employer was dependent on the respondent reporting and interviewer decision-making. However, innovations in the CAPI instrument have allowed the CPS employer to be automatically identified. This is accomplished by sorting the list of employers from most recent to least recent, by stop date. In the unusual event that multiple employers have the same stop date, respondents are asked about the employer for whom they worked the most hours (last week/during the last week they worked), and the interviewer chooses the appropriate employer from a roster of potential CPS employers. If a unique CPS employer is identified (only one employer has the most recent stop date), the interviewer is presented immediately with a roster of employers to verify for accuracy with the respondent. The CPS employer appears at the top of the roster. Because the

presence of the name of a first employer indicates that the respondent has worked since the date of last interview and that a CPS employer exists, there is no need to explicitly ask the respondent or to have the interviewer review information and enter an answer. Instead the answers to these questions are machine generated, after checking for the presence of a first employer name.

This results in some differences in appearance for these variables, compared to earlier PAPI rounds. Noticeable examples of these differences include R(41774.)(Q6-44), R(41818.)(Q6-48) and R(41819.)(Q6-49).

Possible Change in Order of Reporting and/or Repetition of Children (in Verification of Children's Record Form)

Users of the Fertility data are familiar with a series of questions asked in the even-numbered survey years since 1986 in which the information on the Biological and Non-Biological Children's Record Form (CRF) is verified (including name, birthdate, gender and other elements). In the PAPI survey, the interviewer would verify the information for each child and correct any information necessary for that child, marking the information changed in the appropriate column. Properly administered, these verifications would have been made for each child in sequence, from first listed on the CRF to last. The CAPI series that approximates this verification series is structured in the same way. There are however, several differences in the way information is entered and in the possible **order** in which children are verified.

The interviewer is presented with the complete roster of children (biological or non-biological). On the first line of this roster is a line reading "ALL (OTHER) CHILD INFORMATION CORRECT". The interviewer proceeds down the list of children from first to last, verifying the information for each child. When all information is correct for all children listed, s/he chooses the first line reading "ALL (OTHER) CHILDREN CORRECT" as the answer to that question. Instead of a variable for each child indicating whether the information for that child is correct, there is simply one response chosen at the point that the respondent confirms that there are no (more) errors in the information for the children listed.

If the respondent does indicate an error in the information for one or more children, there is no longer a formal order in which the children's information is corrected. For instance, the respondent may have three children listed. Of these there may be errors in the information for children #2 and #3. If the respondent identifies the error in child #3's information first, the interviewer is likely to identify that child first and correct his/her information first. In such a case, the first child identified who is in need of corrected information would be child #3 and the second child identified would be child #2. It is also possible for the respondent to indicate one problem with a child's information, which the interviewer then corrects, and then for the respondent to indicate a second problem with the same child's information, (which the interviewer may correct in a separate loop). In other words, the same child may be identified as the first and second child etc. needing corrected information, with each correction representing a separate piece of information on the roster.

While such patterns are not common, users of this verification data should not automatically assume that the children identified in this series are in the same order as they would be listed on the rosters. Nor should the assumption be made that each loop in this verification series contains a different child.

The final corrected data for all reported children, including corrections made during the verification sequence, can be found in record type CRFBIO for 1993. Reference numbers range between R(44127.) and R(44162.) These variables are the traditional variables included on the *Biological Children's Record Form*.

Children Included in Cyclical Series (Feeding, Child Care)

Unlike the roster verification described above and the pregnancy question series described below, other question sequences pertaining to the children of respondents (particularly female respondents) now include automatic verification of each child for question eligibility. The feeding series and the Child Care Section of the questionnaire (section 10), are two such series of questions. In previous PAPI surveys, only children who were eligible for these series were present in the data. The CAPI instrument however, checks the eligibility of each child on the roster, first through last, to be administered these questions.

This results in differences in the patterns of data from those found in past rounds. Most noticeable will be the fact that each loop of these series will contain information on the corresponding child from the biological child roster. The first loop of the feeding series will contain information on the first child on the roster, the third loop will contain information on the third child, etc. In previous rounds the first loop might contain information on the third child (if that was the first one eligible for the series), etc. In addition, each loop will not necessarily contain substantive information. Only loops for eligible children would contain substantive information.

As noted above, users who wish to attach information on different aspects of female respondents' children to child-specific records should be careful to check the specific child id of children on whom information is collected. This will ensure that the appropriate information is attached to the correct children in any child-based file.

Possible Change in Order of Reporting Children in "Pregnancy" Sequence

Children for whom pre-natal and neo-natal information is required are identified by the interviewer in similar fashion to those needing corrections to roster information. A roster is presented to the interviewer from which s/he chooses the appropriate child to which the "pregnancy" series in the Fertility Section should pertain. The CAPI instrument allows for information to be collected on up to five pregnancies since the last expanded fertility survey. Theoretically, children about whom pregnancy information is required should be identified from oldest to youngest. However, it is possible for the interviewer to enter a younger child first, and then ask questions about an older child. The assumption cannot be made therefore, that the first child for whom pre-natal/neo-natal information is collected is necessarily the first child born since the date of the last expanded fertility interview, and so on. However, the frequency of children being reported out of order should be low, because the instances of female respondents having more than one pregnancy since the last expanded fertility interview is low.

Users attempting to attach information on different aspects of female respondents' children to child-specific records, either independently, or using the Child data sets as a base, should be careful to check the specific child id of children on whom information is collected, in order to ensure that the appropriate information is attached to the correct children.

Institution of Event Histories for Program Reciprocity and Changes in Weeks of Unemployment Reciprocity

In previous PAPI survey rounds, the program reciprocity sequences, (including unemployment compensation for the respondent and respondent's spouse, AFDC, government food stamps and other welfare/public assistance), consisted of a discrete reporting of each source of reciprocity in the months of the calendar year preceding the survey year. In the CAPI instrument, all program reciprocity sequences have been changed to an event history format, beginning with last reported month of receipt, and continuing through the current interview month. Information on between five and six spells of reciprocity from each source is accommodated by the instrument (depending upon whether the respondent was receiving in December of the year before the last interview, the last month possible

for him/her to have reported receipt as of the 1992 interview). The respondent reports the month in which s/he began receiving for each spell, and the month in which s/he stopped receiving for each spell. The average amount received per week/month during each spell was also collected.

One change from the PAPI survey rounds has been in our ability to discern the number of **weeks** during which unemployment compensation was collected. In PAPI survey instruments, a discrete figure for "number of weeks received unemployment compensation" in the calendar year preceding the survey year was collected from the respondent. In the CAPI instrument however, the dates of unemployment compensation receipt were only collected as month and year in which the spell(s) began and ended.

This results in specific week numbers being unavailable for calculating the exact number of weeks during a given calendar year, that a respondent received unemployment compensation. Instead, the beginning month and ending month of a spell are used to determine the number of months in which unemployment was received. Then, the number of months received is multiplied by 4.3 (average number of weeks in a month). This number is used as the total number of weeks received unemployment in a given calendar year. The average **amount** received is an average of the amounts reported during each spell falling within the appropriate calendar year. These are the figures used for number of weeks received unemployment, and average amount received per week by the respondent and his/her spouse, when calculating TOTAL NET FAMILY INCOME.

The possibility of over-estimation of the number of weeks received unemployment compensation exists under this strategy, as it is not always the case that receipt of unemployment during a particular month equates to receipt of unemployment for **all** weeks during that month. Any over-estimation of weeks would probably range from one to seven or eight weeks. These weeks could fall in the beginning and ending months of a spell of unemployment receipt. For a respondent who received unemployment for only one week in the beginning month and one week during the ending month of a spell, the estimation of weeks in that spell could exceed the actual weeks received by approximately seven or eight weeks. CHRR does not have reason to believe that this potential over-estimation of weeks of unemployment receipt is a major source of distortion in the calculation of TOTAL NET FAMILY INCOME.

Number of Program Reciprocity Spells Reported

Each respondent is allowed to report up to five new program reciprocity spells for each type of program. If five is insufficient, the opportunity is provided to report the month and year of most recent receipt. Those reporting that they received in December of the year before the previous interview year (the most recent month in which they could have reported receiving before the 1993 interview) were able to report an ending date for this spell that was technically still open at the date of last interview. This actually constitutes a sixth spell, on which information is collected **prior** to the five new spells allowed since the beginning of the last interview year.

An attempt has been made to reflect these differences in the KWIC titles. However, users should pay close attention to the KWIC titles and the variable content when using these variables. The questionnaire is very helpful in this regard. For those receiving in December of the year before the last interview year, information on the continuation of that spell is generally labelled spell #1, for those having received in December. However, the first of the five new spells allowed is also labelled spell #1. This is actually spell #2 for those who had an "open" spell entering the 1993 interview.

Negative Numbers in Data

In the course of the CAPI interview, negative numbers may result from calculations and be used subsequently in the interview. Two of these variables are present in the data set. These are R(43343.)

and R(43353.) (difference between amount of child support the respondent and/or the respondent's spouse actually received and were supposed to receive in 1992, respectively). In the course of the CAPI survey, some of the values contained in these variables are actually negative values. However in processing the data for release, the negative signs have been dropped from the values. In order to determine if the values of these variables were originally negative values, users can take one of two steps: 1) perform the calculations themselves using the following formulas, R(43340.)-R(43341.) and R(43350.)-R(43351.) (for respondent and spouse respectively); or 2) check the two questions to which respondents with negative numbers would have been skipped, R(43346.) and R(43356.) for respondent and spouse respectively. If R(43346.) or R(43356.) have valid responses, then R(43343.) and R(43353.) respectively originally contained negative numbers which were converted to positive numbers for the purpose of including the figure in the text of the subsequent questions.

Respondent not Included on Household Roster

Information about the respondent has traditionally been collected during the Household Interview, comparable to the information collected on all other household members at that stage of the interview. This information includes: relationship to respondent, age, gender, highest grade completed, and whether the household member worked during the calendar year preceding the interview year. Because all of this information for the respondent is either available elsewhere in the respondent's record and is not subject to change (relationship, age, gender), or is updated during the course of the survey (highest grade completed, working in past calendar year), it was not collected during the 1993 Household Interview as was the case in past rounds. Therefore, the respondent is not present on the household roster. Information on all other members of the household is collected as usual. This does not affect the Family Size variable, which is still computed using the Household Enumeration or Roster. The respondent has been accounted for in the computation of Family Size. Researchers computing independent family size measures and wishing to include the respondent should remember to initialize the variable to "1".

Initial Versus Final Versions of Information Sheet Variables

During the course of running the CAPI survey, certain information about the respondent as of the most recent interview is employed. Often the respondent is given an opportunity to verify this existing information before the interviewer proceeds to collect updated information. Many of the information sheet variables (contained in the record type LASTINFO) have either the notation "(INITIAL)" or "(FINAL)" appended to the KWIC TITLE. A variable characterized as "(INITIAL)" reflects the value of that piece of data from the previous interview BEFORE the respondent was given a chance to verify its accuracy. A variable characterized as "(FINAL)" reflects the value of that piece of data AFTER the respondent has had a chance to verify, and possibly change its value. While the majority of cases will contain the same values on both the "initial" and "final" versions of these data, there are also a number of cases where those values are different. These are the cases where the respondent is disputing the initial value and has corrected or amended the information from the previous interview.

Introduction to 1994 CAPI Questionnaire and Codebook

The 1994 questionnaire and codebook continues many of the conventions established and discussed in the first part of this appendix, *Introduction to 1993 CAPI Questionnaire and Codebook*. And whenever possible, comparability has been maintained between the 1994 data and documentation and that of previous survey years. However, there have been some significant changes, additions and improvements between the 1993 and 1994 data releases. The following is a discussion of some of these significant differences.

I. NEW DOCUMENTATION ITEMS AND CHANGES IN EXISTING DOCUMENTATION

Questionnaire

The 1994 printed questionnaire has been simplified. The essential elements for reading and following the flow of the questionnaire have been distilled. As a result, the 1994 questionnaire resembles much more closely the paper-and-pencil instruments that accompanied the 1979-1992 data releases. See the 1994 questionnaire and accompanying documents for further discussion of format and contents.

New appendices

Two new appendices have been added to the set of documentation called the *Codebook Supplement*. The first is entitled *Appendix 16: The 1994 Recall Experiment*. The second is named *Appendix 15: Program Reciprocity*. The first new document discusses an experiment that was conducted with the 1994 respondents to gain a better understanding of the effects of a biennial survey on respondent recall of life events. The second appendix seeks to clarify the question modules devoted to program reciprocity event histories (receipt of unemployment benefits, AFDC, Food Stamps and other welfare benefits), first implemented in the 1993 NLSY79 survey and continued in the 1994 wave.

II. CHANGES IN CODING CONVENTIONS

Class of Worker variables

In 1994, the coding of the class of worker variables for each employer changed for the first time in the 16 waves of the NLSY79. This change resulted from the emulation of the substantially revised CAPI version of the actual *Current Population Survey*, in several modules of the NLSY79 (see discussion below). A category for “non-profit organization (including charitable)” was added. The response categories prior to 1994 were as follows:

- 1 An employee of a PRIVATE company
- 2 A GOVERNMENT employee
- 3 Self-Employed in OWN business
- 4 Working WITHOUT PAY in a family business or farm

The response categories implemented in 1994 are as follows:

- 1 Government
- 2 Private for profit company
- 3 Non-profit organization (including charitable)
- 4 Self-employed
- 5 Working in a family business

Users may recode Class of Worker variables from previous years and/or create composite variables to achieve relative comparability in this set of variables between the 1979-1993 variables and the 1994 variables (and those in future years). See the discussion below for further information on the manner in which class of worker questions are administered in the 1994 survey.

Employment Status Recode (ESR) variables

Beginning in 1994, the substantive meaning of some of the codes assigned to the Employment Status Recode variables has changed and/or new substantive categories have been added. This was necessitated by the revision of the CPS section, modeled on the actual *Current Population Survey* (see discussion below). The Employment Status Recode variables were in turn modified to emulate the “Monthly Labor Status” variable computed by the Census Bureau from data generated from the *Current Population Survey*. Response codes prior to the 1994 were as follows:

ESR

- 1 WORKING
- 2 WITH A JOB NOT AT WORK
- 3 UNEMPLOYED
- 4 KEEPING HOUSE
- 5 GOING TO SCHOOL
- 6 UNABLE TO WORK
- 7 OTHER
- 8 IN ACTIVE FORCES

ESR (collapsed)

- 1 EMPLOYED
- 2 UNEMPLOYED
- 3 OUT OF THE LABOR FORCE
- 4 IN ACTIVE FORCES

The response categories implemented in 1994 are as follows:

ESR

- 1 EMPLOYED
- 2 EMPLOYED - ABSENT FROM JOB
- 3 UNEMPLOYED - ON LAYOFF
- 4 UNEMPLOYED - LOOKING FOR WORK
- 5 NOT IN LABOR FORCE - RETIRED
- 6 DISABLED
- 7 NOT IN LABOR FORCE - OTHER

The categories for the collapsed version of ESR in 1994 are the same as in previous years.

Revised 1984-1994 FICE Code and 1979-1993 Highest Grade Completed and Enrollment Status variables

Revised versions of the created Highest Grade Completed and Enrollment Status variables from 1979-1994 have been added to the NLSY79 main data file. In addition revised FICE code variables for 1984-1994 have been added to the NLSY79 Geocode data file. For each school identification (FICE code) variable, a variable containing a special edit code was added to the Geocode data file as well. This data item identifies which of several possible types of codes were assigned to a given institution.

The original (unrevised) variables for Highest Grade Completed, Enrollment Status and FICE code remain in the data files, along with the revised variables.

FICE Code Data

An examination by CHRR personnel of the NLSY79 FICE code data between 1984 and 1994 resulted in the identification and correction of the following data problems:

- ◆ data entry error and other inaccuracies;
- ◆ multiple FICE codes assigned to the same university;
- ◆ institutions lacking FICE codes;
- ◆ international universities;
- ◆ unidentified institutions
- ◆ incorrectly identified invalid skips.

Revisions to FICE code data was made in approximately 1250 cases. The FICE code data and accompanying special edit code variables are available only to those who satisfactorily complete the Geocode licensing procedure. For further details concerning the revisions to the FICE code data, see *Attachment 105: Addendum to FICE CODES*, which is part of the Geocode documentation package.

Highest Grade Completed and Enrollment Status

Two main sources of error were identified in the Highest Grade Completed (HGC) variables from 1979-1994:

- ◆ grade “reversals”, in which a respondent completed a lower grade in a later year, rather than staying in the same grade or advancing;
- ◆ respondents with incomplete or ambiguous school information preventing the computation in a given year of a HGC or Enrollment Status (a problem which can then carry through subsequent years).

Examination of the longitudinal record resulted in one or more changes to the created Highest Grade Completed as of May 1st Survey Year variables from 1979-1994 for approximately 3500 respondents. Created enrollment status variables were revised where necessary, based on the revised HGC variables. See *Appendix 8: Highest Grade Completed and Enrollment Status Variable Creation 1990-1994* for details.

III. CHANGES IN PLACEMENT OF DATA AND/OR CODEBOOK ORDER

1980 CPS Occupation and Industry Codes

In the 1994 CAPI data, the 1980 Occupation and Industry codes for the CPS (current or most recent) employer are once again contained in the Employer Supplement. However unlike 1993, the variables for the 1980 Census codes are once again separate from those containing the 1970 occupation and industry codes for non-CPS employers in Employer Supplement # 1. See the discussion below for further information on the manner in which occupation and industry questions were administered in the 1994 survey.

Employer ID from Previous Survey Year

In the 1993 data file, the variables containing the employer IDs from the previous survey year were placed directly preceding the usual earnings questions in the appropriate Employer Supplement. In 1994, these variables have been returned to a position near the beginning of each employer supplement, similar to the 1979-1992 survey years.

CPS and On Jobs sections

In the NLSY79 1993 CAPI data file, the positions of the CPS and On Jobs (employer inventory) sections were changed from their traditional order in the 1979-1992 waves. The 1993 On Jobs section was administered first, directly preceding the CPS section. This allowed the current/most recent (CPS) employer to be determined, (a process which was automated in 1993), prior to administering the CPS section. The goal was to eliminate the error in collection of information on specific employers in the CPS section, and the variation in the order in which the Employer Supplements are administered.

Between 1979 and 1993, certain employer-specific questions pertaining to the CPS employer were contained in the CPS section itself, while similar or identical questions for all other employers were asked during the Employer Supplements. However, in 1994, all questions relating to specific employers were transferred to the Employer Supplements. This eliminated the need for the CPS employer to be established prior to the CPS section was administered. Therefore the order of the CPS and On Jobs sections have been shifted back to the 1979-1992 pattern, with the CPS section being administered before the On Jobs section.

Class of Worker variables for new employers are collected in two questions, depending upon whether business ownership is reported by someone in the respondent's household.

IV. CHANGES IN DATA COLLECTION PROCEDURES WHICH AFFECT FORM AND/OR APPEARANCE OF NLSY79 1994 DATA**NLSY79 Emulation of the Revised CAPI *Current Population Survey***

Several segments of the 1994 survey were modified significantly in an effort to emulate as closely as possible related segments of the actual *Current Population Survey*. Specifically, these sections include the CPS section itself, questions on usual earnings with each employer and the segments which collect occupation, industry and class of worker for each employer.

CPS Section

The CPS section has been revised substantially. It continues to be modeled upon the *Current Population Survey's* section on activity in the last week and last four weeks, for which a conversion to a CAPI administration began in 1994. While essentially the same information is collected as in past years, it is somewhat more extensive and allows a more well-defined labor force status to be identified. In addition, specific information about the current or most recent employer (CPS employer) has been completely removed from the CPS section and placed in the first Employer Supplement. The current NLSY79 CPS section focuses only on general labor force status in specified time frames, not on any individual employer.

Usual Earnings

The questions concerning usual earnings of respondent and spouse/partner have also been revised based upon the *Current Population Survey*. In the NLSY79 1979-1993 surveys this information was collected in two questions, one for the actual rate of pay and one for the time unit for that rate of pay. Beginning with the 1994 NLSY79 survey, this series of questions has been expanded significantly. The series now allows more specificity in handling a given time unit for the rate of pay and includes the following elements:

- ◆ For hourly rates of pay:
 - Separate reporting of overtime pay if relevant and hours per week worked at the overtime rate;
 - Reporting and/or verification of hours worked per week at the usual hourly rate of pay;

- Calculation of usual weekly wage taking overtime rate of pay and hours per week at both overtime and usual hourly rate of pay;
- Opportunity to correct specific data used in the calculation of usual weekly earnings and recalculate usual weekly earnings based upon corrected figures;
- ◆ For annual rates of pay:
 - The number of weeks per year for which the respondent was paid;
- ◆ For non-hourly rates of pay:
 - Inclusion or non-inclusion of overtime amount in usual earning report for time period as appropriate, depending upon whether respondent said that s/he received any overtime pay;
- ◆ For all rates of pay (hourly and non-hourly):
 - Bounding of figures reported for usual earnings, based either on the reported time period for the rate of pay or on the respondent's reports of usual earnings on previous interview dates with the same employer;
 - Several opportunities for verification and/or correction of rates of pay that fall outside a certain range or that the respondent reports were recorded or calculated inaccurately.

Occupation, Industry and Class of Worker

As with the usual earnings information and the CPS section itself, the series of questions soliciting information on occupation, industry and class of worker for each employer was revised to resemble those used in the actual *Current Population Survey*. The collection of this data was done in a manner relatively similar to that in the 1979-1993 surveys. The questions soliciting the descriptions of occupation, activities and duties and industry resembled those in past survey rounds. However, a pattern of verification of past information was adopted in the 1994 survey. If the employer is one reported during the previous interview, the description of the position given at the last interview of the occupation and activities and duties is read to the respondent. The respondent then confirms that the existing description is still correct, or says that it is not correct. If the existing description is reported to be incorrect, a new, updated and/or augmented description of the occupation and activities and duties is given. The industry and class of worker information is not recollected for pre-existing jobs, regardless of whether the occupation is changed or updated. If the respondent confirms that the description of the occupation is accurate from the previous interview, no new information is collected. The codes from the past interview for occupation, industry and class of worker are all retained in the 1994 data. For a newly reported employer, not present at the last interview, information is collected for all three types of data as appropriate, depending upon employer characteristics (number of hours worked per week, number of weeks worked with employer since the date of last interview/start date).

Self-administered Drug Use Supplement

The 1994 NLSY79 CAPI instrument included a confidential self-administered Drug Use Supplement, resembling closely in content that included in the 1992 PAPI instrument. However, as with the rest of the survey, the Drug Supplement was administered as part of the electronic instrument, directly following the Income and Assets section of the questionnaire. When the interview reached the beginning of the Drug Use Supplement, the interviewer turned the laptop to the respondent and the respondent was asked to follow through the introductory instructional screens and then the actual Drug Use Supplement, choosing the responses him/herself.

With PAPI versions of the self-administered Drug Use Supplements, respondent confidentiality was maintained by supplying him/her with an envelope into which s/he would place the Drug Use Supplement after s/he was finished and seal. Field interviewers were not permitted to review the

module. This data was connected to the respondent's entire record only after it reached the NORC central office staff and was data entered.

The CAPI version of the Drug Use Supplement employed an "electronic envelope" of sorts, in order to preserve the same level of confidentiality for the respondent. Once the respondent finished answering the questions and exited the range of questions within the Drug Use Supplement, that module was automatically hidden from view. This prevented interviewers in the field from reviewing the Drug Use Supplement and respondents' individual answers. The data could only be read once it was transmitted to the NORC central office and processed.

Introduction to 1996 CAPI Questionnaire and Codebook

I. CHANGES IN EXISTING DOCUMENTATION

Glossary of Save Array Names

The section entitled *Glossary of Save Array Names* in the 1993 and 1994 Questionnaire documents has been eliminated in the 1996 Questionnaire. In the 1993 and 1994 questionnaires, the names of save arrays (data locations) appeared in the actual question text. Users could then look up the definitions for the data in these save arrays in the *Glossary of Save Array Names*. In 1996, the save array definitions have been inserted directly into the text of the questions in place of the save arrays instead. (See previous section titled “Introduction to 1993 CAPI Questionnaire and Codebook” in this appendix for further description of the term “save array”.)

II. CHANGES IN APPEARANCE OR PRESENTATION OF DATA

“Consolidated” Variables

The 1993 and 1994 data files contained “consolidated” variables. These were existing variables in which data from other variables had been combined, to produce one more inclusive data item. These variables intended to duplicate single data items as they existed in the Paper-and-Pencil data files prior to 1993. In the 1996 data file, no actual survey data points have been used to consolidate data from other survey variables. Instead, where appropriate, new variables have been created, and data from several questions consolidated into those created variables. These items include variables on marital status changes, payrates for the respondent and a spouse/partner if applicable, dates of military enlistment and reasons for within-job gaps.

Reciprocity History Variables

The 1979-1996 NLSY79 release includes a large series of variables pertaining to the history of program reciprocity for unemployment, AFDC, government food stamps and SSI/other public assistance. Variables containing information on amounts received month-by-month from January, 1978, and the source of data for each month, can be found in the record type RECIP_MON. Variables summarizing information on annual program receipt can be found in record type RECIP_YR.

The purpose of these variables is to provide users with a concentrated group of variables from which summary statistics on program receipt can more easily be constructed. For more information on this new series of variables, see *Appendix 15: Reciprocity Event Histories*.

Introduction to 1998 CAPI Questionnaire and Codebook

I. CHANGES IN EXISTING DOCUMENTATION

Area of Interest

In previous rounds, the field Record Type on the data CD contained the basic topic of the data points. In this round, the title “Area of Interest” is now used in the same fashion and replaces the field Record Type.

II. CHANGES IN APPEARANCE OR PRESENTATION OF DATA

Introduction to the Use of the Tilda (~) in the Question Names

In previous rounds, question names were separated from additional information, such loop numbers, with an underscore. Beginning in 1998, the question names are separated from that other information with a tilda (~). The question types affected are discussed below.

Presentation of Dates in Question Names

Questions which contain date data are broken down into month, day, and year entries in the codebook. The question name is separated with a tilda and the time unit is represented with an M, D, or Y. For example:

What month and year was that first diagnosed?

R63562.00	H40-CHRC-1A~M	This will pull the month
R63562.01	H40-CHRC-1A~Y	This will pull the year

Loops

In previous CAPI rounds, the fact that a question was part of a loop was indicated with one decimal and a loop number after the question name. In 1998, the convention has changed slightly with loop numbers now being indicated with **two** decimals and the loop number. For example:

Could you have returned to work last week if you had been recalled?

R58467.00	Q5-51.01	This will pull loop number one.
R58468.00	Q5-51.02	This will pull loop number two.

Select all questions / Multiple Fields

Several questions allow the respondent to select many answers from an answer list or enter several pieces of information. In previous CAPI years, these answers were displayed with an underscore separating each answer choice based on the sequence. In 1998, each answer choice receives it's own entry in the codebook, with the number reflecting the value of the answer choice in the pick list. For example:

Types of compensation based on performance

R60566.00	QES-PAYMT60A.01~000001	Answer w/ value 1 in loop 1
R60566.01	QES-PAYMT60A.01~000002	Answer w/ value 2 in loop 1
R60566.02	QES-PAYMT60A.01~000003	Answer w/ value 3 in loop 1
R60566.03	QES-PAYMT60A.01~000004	Answer w/ value 4 in loop 1
R60566.04	QES-PAYMT60A.01~000005	Answer w/ value 5 in loop 1

R60566.05	QES-PAYMT60A.01~000006	Answer w/ value 6 in loop 1
-----------	------------------------	-----------------------------

